

massachusetts institute of technology — artificial intelligence laboratory

---

# Inferring 3D Structure with a Statistical Image-Based Shape Model

Kristen Grauman, Gregory Shakhnarovich  
and Trevor Darrell

AI Memo 2003-008

April 2003

## Abstract

*We present an image-based approach to infer 3D structure parameters using a probabilistic “shape+structure” model. The 3D shape of a class of objects may be represented by sets of contours from silhouette views simultaneously observed from multiple calibrated cameras. Bayesian reconstructions of new shapes can then be estimated using a prior density constructed with a mixture model and probabilistic principal components analysis. We augment the shape model to incorporate structural features of interest; novel examples with missing structure parameters may then be reconstructed to obtain estimates of these parameters. Model matching and parameter inference are done entirely in the image domain and require no explicit 3D construction. Our shape model enables accurate estimation of structure despite segmentation errors or missing views in the input silhouettes, and works even with only a single input view. Using a dataset of thousands of pedestrian images generated from a synthetic model, we can perform accurate inference of the 3D locations of 19 joints on the body based on observed silhouette contours from real images.*

This work was supported by the Department of Energy Computational Science Graduate Fellowship (CSGF) and the DARPA Human Identification at a Distance (HID) program.

# 1. Introduction

Estimating model shape or structure parameters from one or more input views is an important computer vision problem. Classic techniques attempt to detect and align 3D model instances within the image views, but high-dimensional models or models without well-defined features may make this type of search computationally prohibitive. Rather than fit explicit 3D models to input images, we explore reconstruction and parameter inference using image-based shape models which can be matched directly to observed features. We learn an implicit, image-based representation of a known 3D shape, match it to input images using a statistical model, and infer 3D parameters from the matched model.

Implicit representations of 3D shape can be formed using models of observed feature locations in multiple views. With sufficient training data of objects of a known class, a statistical multi-view appearance model can represent the most likely shapes in that class. Such a model can be used to reduce noise in observed images, or to fill in missing data.

In this paper we present an image-based approach to infer 3D structure parameters. A probabilistic “shape+structure” model is formed using a probability density of multi-view silhouette images augmented with known 3D structure parameters. We combine this with a model of the observation uncertainty of the silhouettes seen in each camera to compute a Bayesian estimate of structure parameters. A reconstruction of an observed object yields the multi-view contours and their 3D structure parameters simultaneously. To our knowledge, this is the first work to formulate an image-based statistical shape model for the inference of 3D structure.

We also show how the image-based model can be learned from a known 3D shape model. Using a computer graphics model of articulated human bodies, we render a database of views augmented with the known 3D feature locations (and optionally joint angles, etc.) From this we learn a joint shape and structure model prior, which can be used to find the instance of the model class that is closest to a new input image. One advantage of a synthetic training set is that labeled real data is not required; the synthetic model includes 3D structure parameter labels for each example.

The strength of our approach lies in our use of a probabilistic multi-view shape model which restricts the object shape and its possible structural configurations to those that are most probable given the object class and the current observation. Even when given poorly segmented binary images of the object, the statistical model can infer appropriate structure parameters. Moreover, all computation is done within the image domain, and no model matching or search in 3D space is required.

In our experiments, we demonstrate how our shape+structure model enables accurate estimation of structure parameters despite large segmentation errors or even missing views in the input silhouettes. Since parameter inference with our model

succeeds even with missing views, it is possible to match the model with fewer views than it has been trained on. We also show how configurations that are typically ambiguous in single views are handled well by our multi-view model.

Possible applications of the methods presented include fast approximation of 3D models for virtual reality applications, gesture recognition, pose estimation, or image feature correspondence matching across images.

## 2. Previous Work

In this paper we consider image-based statistical shape models that can be directly matched to observed shape contours. Models which capture the 2D distribution of feature point locations have been shown to be able to describe a wide range of flexible shapes, and they can be directly matched to input images [4]. The authors of [1] developed a single-view model of pedestrian contours, and showed how a linear subspace model formed from principal components analysis could represent and track a wide range of motion [2]. A model appropriate for feature point locations sampled from a contour is also given in [2]. This single-view approach can be extended to 3D by considering multiple simultaneous views of features. Shape models in several views can be separately estimated to match object appearance [5]; this approach was able to learn a mapping between the low-dimensional shape parameters in each view.

With multi-view contours from cameras at known locations, a visual hull can be recovered to model the shape of the observed object [11]. Algorithms for fast rendering of image-based visual hulls, which sidestep any geometric construction, have recently been developed [12]. By forming a statistical model of these multi-view contours, an implicit shape representation that can be used for efficient reconstruction of visual hulls is created [8].

Our model is based on a mixture of Gaussians model, where each component is estimated using principal components analysis (PCA). The use of linear manifolds estimated by PCA to represent an object class, and more generally an appearance model, has been developed by several authors [16, 3, 10]. A probabilistic interpretation of PCA-based manifolds has been introduced by [17, 9] as well as in [13], where it was applied directly to face images. As described below, we rely on the mixture of Gaussian probabilistic principal components analysis (PPCA) formulation of [15] to model prior densities.

The idea of augmenting a PCA-based appearance model with structure parameters and using projection-based reconstruction to fill in the missing values of those parameters in new images was first proposed in [6]. A method that used a mixture of PCA approach to learn a model of single contour shape augmented with 3D structure parameters was presented in [14]; they were able to estimate 3D hand and arm location just from a single silhouette. This system was also able to model contours



observed in two simultaneous views, but separate models were formed for each so no implicit model of 3D shape was formed.

### 3. Bayesian Multi-view Shape Reconstruction

While regularization or Bayesian maximum a posteriori (MAP) estimation of single-view contours has received considerable attention as described above, less attention has been given to multi-view data from several cameras simultaneously observing an object. With multi-view data, a probabilistic model and MAP estimate can be computed on implicit 3D structures. We apply a PPCA-based probability model to form Bayesian estimates of multi-view contours, and show how such a representation can be augmented and used for inferring structure parameters. Our work builds on the shape model introduced in [8], where a multi-view contour density model is derived for the purpose of 3D visual hull reconstruction.

Silhouette shapes are represented as sampled points on closed contours, with the shape vectors for each view concatenated to form a single vector in the input space. That is, with a set of  $n$  contour points  $\mathbf{c}_k$  in each of the  $K$  views,

$$\mathbf{c}_k = (\mathbf{x}_0^k, \mathbf{x}_1^k, \dots, \mathbf{x}_n^k), \quad 0 < k < K, \quad (1)$$

a multi-view observation is defined as

$$\mathbf{c} = (\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_K)^T$$

As described in [8], if the vector of observed contour points of a 3D object resides on a linear manifold, then a multi-view image-based representation of the approximate 3D shape of that object should also lie on a linear manifold, at least for the case of affine cameras. Therefore, the shape vectors may be expressed as a linear combination of the 3D bases.

A technique suitable only for highly constrained shape spaces is to approximate the space with a single linear manifold. For more deformable structures, it is difficult to represent the shape space in this way. For example, with the pedestrian data we will use in the experiments reported below, inputs are expected to vary in two key (nonlinear) ways: the absolute direction in which the pedestrian is walking across the system workspace, and the phase of his walk cycle in that frame.

Thus, following [15, 3], we construct a density model using a mixture of PPCA models that locally model clusters of data in the input space with probabilistic linear manifolds. A single PPCA model is a probability distribution over the observation space for a given latent variable, which for this shape model is the true underlying contours in the multi-view image. Parameters for the  $M$  Gaussian mixture model components are determined for the set of observed data vectors  $\mathbf{c}_n$ ,

$1 \leq n \leq N$ , using an EM algorithm to maximize a single likelihood function

$$L = \sum_{n=1}^N \log \sum_{i=1}^M \pi_i p(\mathbf{c}_n | \mathbf{i}) \quad (2)$$

where  $p(\mathbf{c}_n | i)$  is a single PPCA model, and  $\pi_i$  is the  $i^{th}$  component's mixture proportion. A separate mean vector  $\mu_i$ , principal axes  $\mathbf{W}_i$ , and covariance parameter  $\sigma_i$  is associated with each of the  $M$  components. As this likelihood is maximized, both the appropriate partitioning of the data and the respective principal axes are determined. The mixture of probabilistic linear subspaces constitutes the prior density of the object shape.

We assume there is a normal distribution of camera noise or jitter that affects the observed contour point locations in the input images, and we model this as a multivariate Gaussian with covariance  $\Sigma_o$ . A simple model may use a spherical covariance matrix for  $\Sigma_o$ , where the value is a tunable parameter depending on the amount of regularization desired.

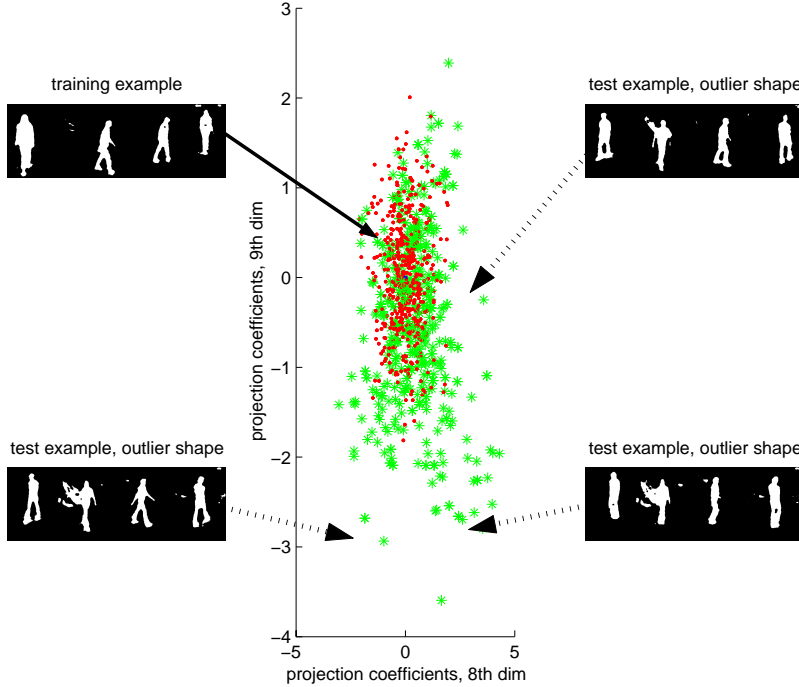


Figure 1: Illustration of prior and observed densities. Center plot shows two projection coefficients in the subspace for training vectors (red dots) and test vectors (green stars), all from real data. The distribution of cleanly segmented silhouettes (such as the multi-view image in top left) is representative of the prior shape density learned from the training set. The test points are poorly segmented silhouettes which represent novel observations. Shown in bottom left and on right are some test points lying far from the center of the prior density. Due to large segmentation errors, they are unlikely samples according to the prior shape model. MAP estimation reconstructs such contours as shapes closer to the prior. Eighth and ninth dimensions are shown here; other dimensions are similar.

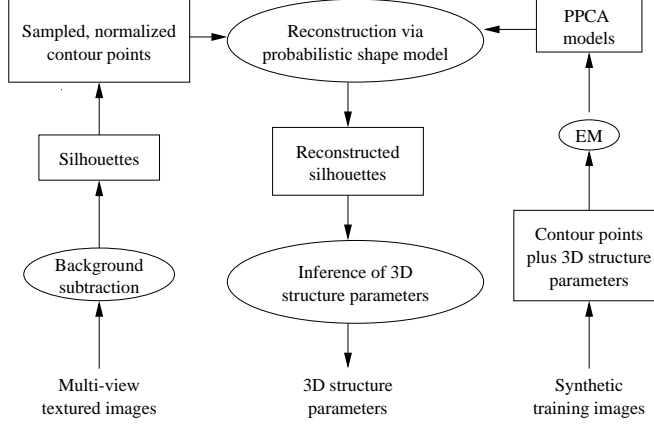


Figure 2: Diagram of data flow in our system.

A MAP estimate of the silhouettes is formed based on the PPCA prior shape model and the Gaussian distributed observation density [15]. The estimate is then backprojected into the multi-view image domain to generate the recovered silhouettes. By characterizing which projections onto the subspace are more likely, the range of possible reconstructions is effectively moderated to be more like those expressed in the training set (see Figure 1).

## 4. Inferring 3D Structure

We extend the shape model described above to incorporate additional structural features within the PPCA framework. A model built to represent the shape of a certain class of objects using multiple contours can be augmented to include information about the object’s orientation in the image, as well as the 3D locations of key points on the object. The mixture model now represents a density over the observation space for the true underlying contours together with their associated 3D structure parameters. Novel examples are matched to the contour-based shape model using the same multi-view reconstruction method described in Section 3 in order to infer their unknown or missing parameters. (See Figure 2 for a diagram of data flow.)

The shape model is trained on a set of vectors that are composed of points from multiple contours from simultaneous views, plus a number of three-dimensional structure parameters,  $\mathbf{s}_j = (s_j^0, s_j^1, s_j^2)$ . Each training input vector  $\mathbf{v}$  is then defined as

$$\mathbf{v} = (\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_K, \mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_z)^T \quad (3)$$

where there are  $z$  3D points for the structure parameters. When presented with a new multi-view contour, we find the MAP estimate of the shape and structure parameters based on only the observable contour data. The training set for this inference task may be comprised of real or synthetic data.

One strength of the proposed approach for the estimation of 3D feature locations is that the silhouettes in the novel inputs need not be cleanly segmented. Since the contours and unknown parameters are reconstructed concurrently, the parameters are essentially inferred from a restricted set of feasible shape reconstructions; they need not be determined by an explicit match to the raw observed silhouettes. Therefore, the probabilistic shape model does not require an expensive segmentation module. A fast simple foreground extraction scheme is sufficient.

As should be expected, our parameter inference method also benefits from the use of multi-view imagery (as opposed to single-view). Multiple views will in many cases overcome the ambiguities that are geometrically inherent in single-view methods.

## 5. Learning a Multi-view Pedestrian Shape Model

A possible weakness of any shape model defined by examples is that the ability to accurately represent the space of realizable shapes will generally depend heavily on the amount of available training data. Moreover, we note that the training set from which the probabilistic shape+structure model is learned must be “clean”; otherwise the model could fit the bias of a particular segmentation algorithm. It must also be labeled with the true values for the 3D features. Collecting a large data set with these properties would be costly in resources and effort, given the state of the art in motion capture and segmentation, and at the end the “ground truth” could still be imprecise. We chose therefore to use realistic synthetic data for training a multi-view pedestrian shape model. We obtained a large training set by using POSER [7] – a commercially available animation software package – which allows us to manipulate realistic humanoid models, position them in the simulated scene, and render textured images or silhouettes from a desired point of view. Our goal is to train the model using this synthetic data, but then use the model for reconstruction and inference tasks with real images.

We generated 20,000 synthetic instances of multi-view input for our system. For each instance, a humanoid model was created with randomly adjusted anatomical shape parameters, and put into a walk-simulating pose, at a random phase of the walking cycle. The orientation of the model was drawn at random as well in order to simulate different walk directions of human subjects in the scene. Then for each camera in the real setup we rendered a snapshot of the model’s silhouette from a

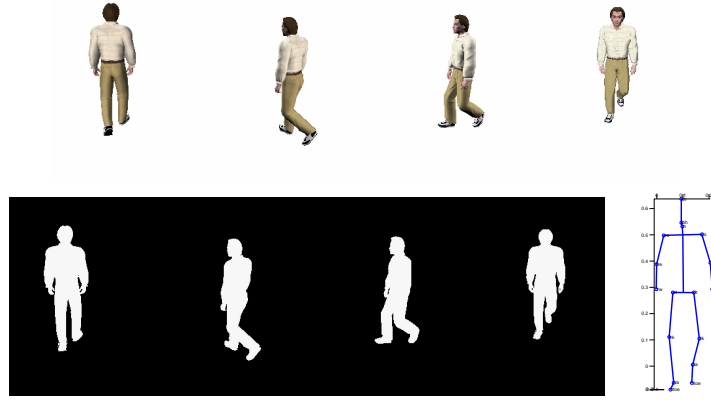


Figure 3: An example of synthetically generated training data. Textured images (top) show rendering of example human model, silhouettes and stick figure (below) show multi-view contours and structure parameters, respectively.

point in the virtual scene approximately corresponding to that camera. In addition to the set of silhouettes, we record the 3D locations of 19 landmarks of the model’s skeleton, corresponding to selected anatomical joints. (See Figure 3.)

For this model, each silhouette is represented as sampled points along the closed contour of the largest connected component extracted from the original binary images. All contour points are normalized to a translation and scale invariant input coordinate system, and each vector of normalized points is resampled to a common vector length using nearest neighbor interpolation. The complete representation is then the vector of concatenated multi-view contour points plus a fixed number of 3D body part locations (see Equation 3).

## 6. Experiments

We have applied our method to datasets of multi-view images of people walking. The goal is to infer the 3D positions of joints on the body given silhouette views from different viewpoints.

For the following experiments, we used an imaging model consisting of four monocular views per frame from cameras located at approximately the same height at known locations about 45 degrees apart. The working space of the system is defined as the intersection of their fields of view (approximately three meters). Images of subjects walking through the space at various directions are captured, and a simple statistical color background model is employed to extract the silhouette foreground from each viewpoint. In the input observation vector for each test example, the 3D pose parameters are set to zero.

Since we do not have ground truth pose parameters for the raw test data, we have tested a separate, large, synthetic

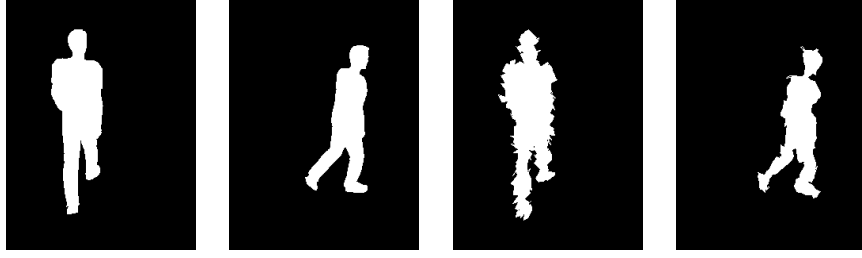


Figure 4: Two left images show clean synthetic silhouettes. Two right images show same silhouettes with noise added to image coordinates of contour points. First has uniform noise; second has nonuniform noise in patches normal to contour.

test set with known pose parameters so that we can obtain error measurements for a variety of experiments. In order to evaluate our system’s robustness to mild changes in the appearance of the object, we generated test sequences in the same manner as the synthetic training set was generated, but with different virtual characters, i.e., different clothing, hair and body proportions. To make the synthetic test set more representative of the real, raw silhouette data, we added noise to the contour point locations. Noise is added uniformly in random directions, or in contiguous regions along the contour in the direction of the 2D surface normal. Such alterations to the contours simulate the real tendency for a simple background subtraction mechanism to produce holes or false extensions along the true contour of the object. (See Figure 4.)

Intuitively, a multi-view framework can discern 3D poses that are inherently ambiguous in single-view images. Our experimental results validate this assumption. We performed parallel tests for the same examples, in one case using our existing multi-view framework, and in the other, using the framework outlined above, only with the model altered to be trained and tested with single views alone. Figure 7 compares the overall error distributions of the single and multi-view frameworks for a test set of 3,000 examples. Errors in both pose and contours are measured for both types of training. Multi-view reconstructions are consistently more accurate than single-view reconstructions. Training the model on multi-view images yields on average 24% better pose inference performance and 16% better contour reconstruction performance than training the model on single-view images.

We have also tested the performance of our multi-view method applied to body pose estimation when only a subset of views is available for reconstruction. A missing view in the shape vector is represented by zeros in the elements corresponding to that view’s resampled contour. Just as unknown 3D locations are inferred for the test images, our method reconstructs the missing contours by inferring the shape seen in that view based on examples where all views are known. (See Figures 5, 6, 8, and 9.)

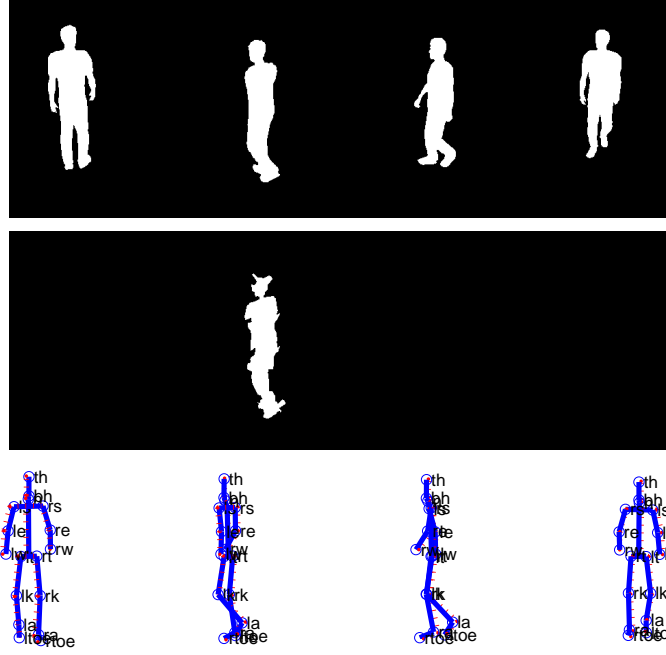


Figure 5: Pose inference from only a single view. Top row shows ground truth silhouettes that are not in the training set. Noise is added to input contour points of second view (middle), and this single view alone is matched to the multi-view shape model in order to infer the 3D joint locations (bottom, solid blue) and compare to ground truth (bottom, dotted red). Abbreviated body part names appear by each joint. This is an example with average pose error of 5 cm.

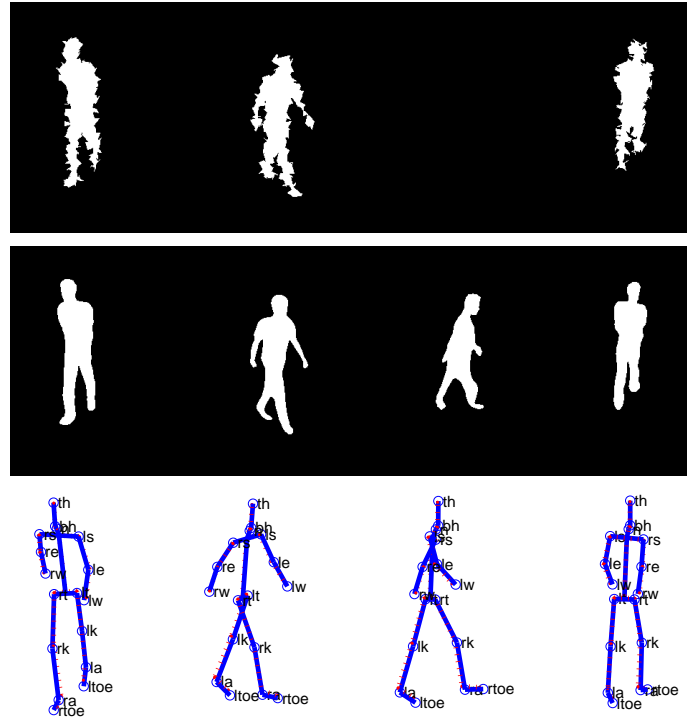


Figure 6: Pose inference with one missing view. Top row shows noisy input silhouettes, middle row shows contour reconstructions, and bottom row shows inferred 3D joint locations (solid blue) and ground truth pose (dotted red). This is an example with average pose error of 2.5 cm per joint and an average Chamfer distance from the true clean silhouettes of 2.3.

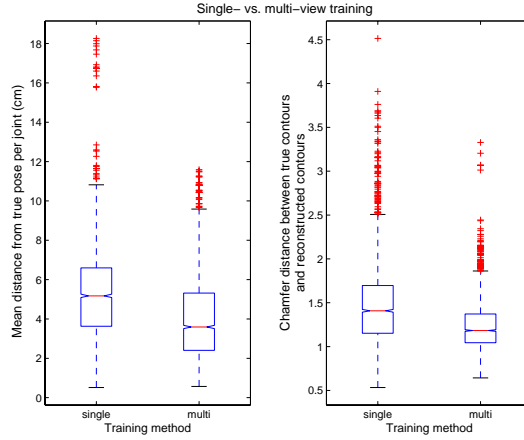


Figure 7: Training on single view vs. training on multiple views. Chart shows error distributions for pose (left) and contour (right). Lines in center of boxes denote median value; top and bottom of boxes denote upper and lower quartile values, respectively. Dashed lines extending from each end of box show extent of rest of the data. Outliers are marked with pluses beyond these lines.

We are interested in knowing how pose estimation performance degrades with each additional missing view, since this will determine how many cameras are necessary for suitable pose estimation should we desire to use fewer cameras than are present in the training set. Once the multi-view model has been learned, it may be used with fewer cameras, assuming that the angle of inclination of the cameras with the ground plane matches that of the cameras with which the model was trained.

Figure 8 shows results for 3,000 test examples that have been reconstructed using all possible numbers of views (1,2,3,4), alternately. For a single missing view, each view is omitted systematically one at a time, making 12,000 total tests. For two or three missing views, omitted views are chosen at random in order to approximately represent all possible combinations of missing views equally. As the number of missing views increases, performance degrades more gracefully for pose inference than for contour reconstruction.

The pose error  $e_f$  for each test frame is defined as the average distance in centimeters between the estimated and true positions of the 19 joints,

$$e_f = \frac{1}{19} \sum_i |e_i|, \quad (4)$$

where  $e_i$  is the individual error for joint  $i$ .

As described above, test silhouettes are corrupted with noise and segmentation errors so that they may be more representative of real, imperfect data, yet still allow us to do a large volume of experiments with ground truth. The “true” underlying



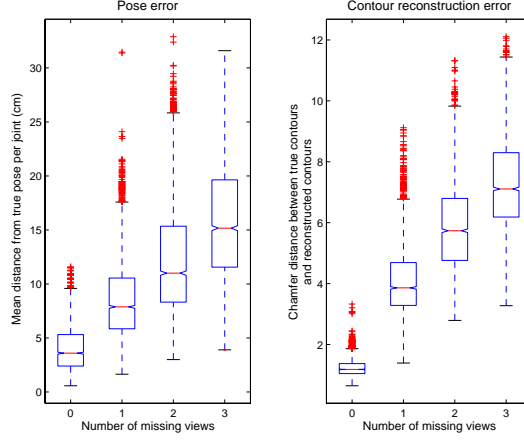


Figure 8: Missing view results. Chart shows distribution of errors for pose (left) and contours (center) when model is trained on four views, but only a subset of views is available for reconstruction. Plotted as in previous figure.

contours from the clean silhouettes (i.e., the novel silhouettes before their contour points were corrupted) are saved for comparison with the reconstructed silhouettes. The contour error for each frame is then the distance between the true underlying contours and their reconstructions.

Contour error is measured using the Chamfer distance. For all pixels with a given feature (usually edges, contours, etc.) in the test image  $\mathbf{I}$ , the Chamfer distance  $\mathbf{D}$  measures the average distance to the nearest feature in the template image  $\mathbf{T}$ .

$$\mathbf{D}(\mathbf{T}, \mathbf{I}) = \frac{1}{N} \sum_{f \in T} d_T(f) \quad (5)$$

where  $N$  is the number of pixels in the template where the feature is present, and  $d_T(f)$  is the distance between feature  $f$  in  $\mathbf{T}$  and the closest feature in  $\mathbf{I}$ .

To interpret the contour error results in Figure 8, consider that the average contour length is 850 pixels, and the pedestrians silhouettes have an average area of 30,000 pixels. If we estimate the normalized error to be the ratio of average pixel distance errors (number of contour pixels multiplied by Chamfer distance) to the area of the figure, then a mean Chamfer distance of 1 represents an approximate overall error of 2.8%, distances of 4 correspond to 11%, etc. Given the large degree of segmentation errors imposed on the test sets, these are acceptable contour errors in the reconstructions, especially since the 3D pose estimates (our end goal) do not suffer proportionally.

Finally, we evaluated our algorithm on a large dataset of real images of pedestrians taken from a database of 4,000 real

multi-view frames. The real camera array is mounted on the ceiling of an indoor lab environment. The external parameters of the virtual cameras in the graphics software that were used for training are roughly the same as the parameters of this real four-camera system. The data contains 27 different pedestrian subjects.

Sample results for the real test dataset are shown in Figure 9. The original textured images, the extracted silhouettes, and the inferred 3D pose are shown. Without having point-wise ground truth for the 3D locations of the body parts, we can best assess the accuracy of the inferred pose by comparing the 3D stick figures to the original textured images. To aid in inspection, the 3D stick figures are rendered from manually selected viewpoints so that they are approximately aligned with the textured images.

In summary, our experiments show how the shape+structure model we have formulated is able to infer 3D structure by matching observed image features directly to the model. Our tests with a large set of noisy, ground-truthed synthetic images offer evidence of the ability of our method to infer 3D parameters from contours, even when inputs have segmentation errors. In the experiments shown in Figure 8, structure inference for body pose estimation is accurate within 3 cm on average. Performance is good even when there are fewer views available than were used during training; with only one input view, pose is still accurate within 15 cm on average, and can be as accurate as within 4 cm. Finally, we have successfully applied our synthetically-trained model to real data and a number of different subjects.

## 7. Conclusions and Future Work

We have developed an image-based approach to infer 3D structure parameters using a probabilistic multi-view shape model. Novel examples with contour information but unknown 3D point locations are matched to the model in order to retrieve estimates for unknown parameters. Model matching and parameter inference are done entirely in the image domain and require no explicit 3D construction from multiple views. We have demonstrated how the use of a class-specific prior on multi-view imagery enables accurate estimation of structure parameters in spite of large segmentation errors or even missing views in the input silhouettes.

In future work we will explore non-parametric density models for inferring structure from shape. We also plan to run experiments using motion capture data so that we may compare real image test results to ground-truth joint angles. In addition, we intend to include dynamics to strengthen our model for the pedestrian walking sequences. We are also interested in how the body pose estimation application might be utilized in some higher-level gesture or gait recognition system.

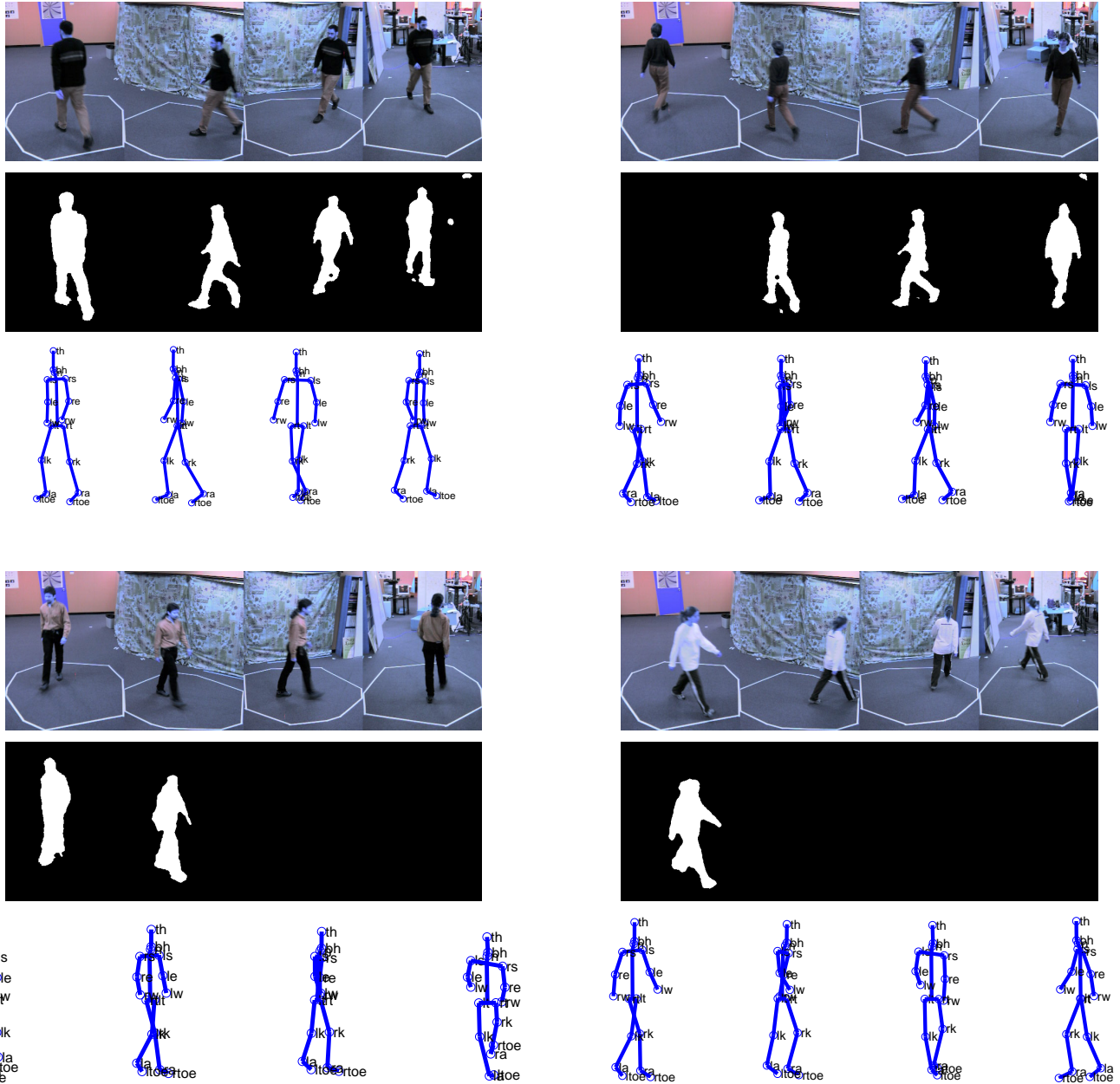


Figure 9: Inferring structure on real data. For each example, top row shows original textured multi-view image, middle row shows extracted input silhouettes where the views that are not used in reconstruction are omitted, and bottom row shows inferred joint locations with stick figures rendered at different viewpoints. To aid in inspection, the 3D stick figures are rendered from manually selected viewpoints that were chosen so that they are approximately aligned with the textured images. In general, estimation is accurate and agrees with the perceived body configuration. An example of an error in estimation is shown in the top left example’s left elbow, which appears to be incorrectly estimated as bent.

## References

- [1] A. Baumberg and D. Hogg. Learning flexible models from image sequences. In *Proceedings of European Conference on Computer Vision*, 1994.
- [2] A. Baumberg and D. Hogg. An adaptive eigenshape model. In *British Machine Vision Conference*, pages 87–96, Birmingham, September 1995.
- [3] T. Cootes and C. Taylor. A mixture model for representing shape variation. In *British Machine Vision Conference*, 1997.
- [4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.
- [5] T.F. Cootes, G.V. Wheeler, K.N. Walker, and C.J. Taylor. View-based active appearance models. *Image and Vision Computing*, 20:657–664, 2002.
- [6] M. Covell. Eigen-points: Control-point location using principal component analysis. In *Proceedings of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Killington, October 1996.
- [7] Egisys Co. Curious Labs. Poser 5 : The ultimate 3D character solution. 2002.
- [8] K. Grauman, G. Shakhnarovich, and T. Darrell. An image-based approach to bayesian visual hull reconstruction. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, Madison, 2003.
- [9] J. Haslam, C. Taylor, and T. Cootes. A probabilistic fitness measure for deformable template models. In *British Machine Vision Conference*, pages 33–42, York, England, September 1994.
- [10] M. Jones and T. Poggio. Multidimensional morphable models. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 683–688, New Delhi, January 1998.
- [11] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, February 1994.
- [12] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan. Image-based visual hulls. In *Proceedings of the 27th Conference on Computer Graphics and Interactive Techniques*, Annual Conference Series, pages 369–374, 2000.
- [13] B. Moghaddam. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):780–788, June 2002.
- [14] E. Ong and S. Gong. The dynamics of linear combinations:tracking 3D skeletons of human subjects. *Image and Vision Computing*, 20:397–414, 2002.

- [15] M. Tipping and C. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.
- [16] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 586–590, Hawai, June 1992.
- [17] Y. Wang and L. H. Staib. Boundary finding with prior shape and smoothness models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):738–743, 2000.